

DOCUMENT RESUME

ED 061 264

TM 001 168

AUTHOR Gentile, J. Ronald
TITLE Toward Excellence in Teaching: Grading Practices.
PUB DATE 71
NOTE 8p.

EDRS PRICE MF-\$0.65 HC-\$3.29
DESCRIPTORS Behavioral Objectives; Course Objectives; *Criterion
Referenced Tests; Educational Accountability;
Effective Teaching; Feedback; *Grades (Scholastic);
*Grading; Measurement Techniques; *Norm Referenced
Tests; Student Evaluation; Student Motivation;
*Teaching Styles; Test Reliability; Test Validity

ABSTRACT

A comparison is drawn between norm-referenced (or competitive) and criterion-referenced testing procedures as used for the purpose of assessing and grading knowledge gained from teaching. It is argued that competitive or norm-referenced grading practices are sadistic, unethical, statistically unsound, and irrelevant to course objectives. Criterion-referenced procedures are advocated as alternatives which avoid the problems of norm-referenced testing.
(Author)

TOWARD EXCELLENCE IN TEACHING: GRADING PRACTICES¹

J. Ronald Gentile

Department of Educational Psychology

State University of New York at Buffalo

Periodically, it pays educators to question their most traditional practices in the hope that such questioning will lead to improvements where needed, or at least to force us to provide rationales for our practices and thus keep us from the educational "mindlessness" which Charles E. Silberman convincingly documents in Crisis in the Classroom. With this goal in mind, I would like to present a point of view on one of the integral parts of teaching as we now know it -- the practice of grading.

An analysis of grading practices is critical for at least two reasons: first, because much of the way students spend their time learning in a given course is determined by the grading procedure of the instructor; and second, because the way in which grading is presently handled, probably by upwards of 90% of the faculty at every educational institution in the country, is atrocious.

The traditional and most widely used grading system is one in which the instructor evaluates the students by differentially ranking or grading them on the basis of their differential performances, usually on a test or paper. Thus it is not unusual for the instructor to give midterm and final examinations; to combine the two scores for each student in some way to arrive at a final course average; and, on the basis of the distribution of scores on this final average, to decide which students should receive grades of A,B,C, or whatever. Such a

¹ This article is based on a viewpoint expressed in the Reporter, State University of New York at Buffalo, March 25, 1971, Vol. 2, No. 26, p.4.

ED 061264

TM 001 168

procedure is technically known as norm-referenced testing since any single person's score is interpreted in terms of the scores of the other students -- the norm. I prefer the term competitive testing for this procedure since one student is forced to compete with another in order to be stamped with the instructor's and society's seal of approval.

To be blunt about it, competitive grading practices are sadistic, unethical, statistically unsound, and irrelevant to the course objectives. Competitive testing is, first of all, sadistic because it pits one student against another, setting up a situation in which cooperation among students is risky. It encourages cheating whenever the probability of being caught is low (which is usually always when the students are ingenious, which is also usually always), and it probably contributes to the loss of library books and journal pages in universities and to writing on the table tops and coded messages in the public schools. Some people might prefer to think of these evidences of competition as providing increased motivation to study hard. I am convinced that they increase motivation to find ways to beat the system, but I doubt that they result in increases in learning course concepts. Even if competitive testing did increase motivation, it is still sadistic because it is based on the assumption that everyone cannot or should not succeed in the course -- that is, achieve the agreed upon objectives. Thus the instructor is telling each student who does not receive the highest grade that he is not as good as the other students.

This is also where the ethics of the procedure enters, though it is more apparent if you think of the effects of competitive testing at the elementary school level. In a society in which the school's function is to provide each child with basic skills necessary for him to select his own pursuit of happiness, each student has a right to succeed in a course. In elementary school, society

expects each child to succeed in various skills such as the three R's, cooperative play, etc. In college, students pay for the privilege of gaining knowledge, skills, or degrees. At either end of the continuum the end result should be the student's success in these goals. In a competitive system, however, any time one student succeeds, at least one other student fails. Teachers do not have the right to play God and decide who should and should not be successes in life. Even were they competent to make such judgments, making them is not part of their job, which is to help and encourage each student achieve the objectives of the course. More on this later.

The third point about competitive testing has to do with the grades of students which are derived from the test scores. Distributions of test scores have been the subject of a great amount of research by scholars in the discipline of educational and psychological measurement. Many concepts could be invoked from this field to support an argument that most tests constructed by instructors have major flaws, one of the most serious being that they provide unreliable measures. However, it is not necessary to argue the statistical unsoundness of competitive grading practices on such esoteric points. Even if the instructor's test were technically sound and perfectly reliable -- which no test is -- any differential grading or ranking of students on the basis of the test scores would still be an arbitrary process. It is arbitrary whether the scores are raw, converted, normalized, standardized, or otherwise. It is arbitrary because at almost any cutoff point the score which falls into the area receiving one grade is not significantly different from the closest score falling into the area receiving an adjacent grade. The finer the discriminations in grading, the more arbitrary the process. Thus, if an instructor decides that a score of 90% correct will earn an A, but 89% will earn a B, he has arbitrarily made this decision, since

in point of fact there is no reliable difference between a score of 89 and 90. In fact, for most tests there is probably no significant difference between scores of 80 and 90.

In a statistical sense, scores which are not significantly different from one another should be considered to be equivalent, and students who receive those scores should be considered to have learned the material equally well. In case we have some nonbelievers on this point, I invite any statistical expert to describe the conditions under which adjacent scores receiving different grades could be considered to be significantly different, and thus provide us with a rational, non-arbitrary procedure for competitive grading. I predict that, in practice, the distribution of scores necessary to sustain the obtained rational procedure will never occur.

All of the above reasons are sufficient in and of themselves to warrant the immediate (and even retroactive, if it were possible) cessation of competitive grading practices. Nevertheless, to add insult to injury, the teacher who engages in such practices is not even doing his job: that is, the whole practice of testing and grading competitively is irrelevant to the process of teaching.

Teaching is an activity which cannot be divorced from learning. To paraphrase Dewey, we should laugh at a salesman who said he sold many items when no one bought any. The analogy to learning is perfect: if any student has not learned, the teacher has not taught him. It is irrelevant to that student if other students have learned. We can say that the teacher was successful in the case of those who learned, but we must also say that he failed in the case of those who did not learn. Of what use, then, is a rank ordering of students from best to worst in performance? What the instructor needs to determine is those students who have achieved the course objectives and those who have not.

For those who have not, the instructor needs to teach them so that they achieve the objectives. Why else are faculty members being paid to teach?

For those who would argue that businesses, industry, and graduate schools depend upon the class ranks of students, I argue that is too bad. It is not the teacher's job, but the selector's (employer or school), to devise ways to select students for their positions. The sooner the schools get out of the business of maintaining competitive permanent record files (which seem indestructible and available to almost anyone but the student), the better.

Having presented my case against traditional grading procedures, I want to be careful not to place myself on the side of those who have reacted so strongly against the evils of these procedures to have gone to the opposite extreme -- the extreme of no assessment of learning. By the argument in the last few paragraphs it can be seen that assessment of what has been learned is an integral part of teaching. Without it, teaching cannot be claimed to have occurred. Thus, some assessment of whether each student has attained the objectives is necessary, although tests are not the only way of assessing this. Too many, especially in some of the free school movements, have abandoned assessment entirely. While this is a neat solution to the problems of sadism, unethicity and statistical unsoundness which have been raised, and often has the added attractions of letting students participate in the establishment of objectives, we cannot claim that teaching has occurred until assessment of the learning reveals that the objectives have been met.

Some may argue that learning has occurred even if the objectives have not been attained, and they are probably correct. However, learning is a continual process and does not need a teacher. Thus, if you want to justify yourself as a teacher, you must demonstrate that what you taught to a given student was learned by him.

If neither the competitive assessment nor nonassessment approaches to grading is appropriate to excellent teaching, what is? As I see it, the answer lies in what is called criterion-referenced or mastery testing. The procedure involved is so named because each student is evaluated in relation to the course objectives taken as the criterion, and he must demonstrate that he has mastered these objectives. This means that each student is judged solely in relation to these pre-established criteria (which again could have been determined by the student, the instructor, or both) and independently of the performance of any other student. It also means that until a student has learned the material, he is merely in the process of learning. Students who take longer than others to learn should not be stigmatized, but should be helped to learn. (Often it is possible to have other students assist them since cooperation will not harm anyone's class rank.) Individual differences in rate of learning, of course, will still exist, but from the standpoint of the instructor they are unrelated to his purpose -- namely, to have each student achieve the course objectives.

The perceptive reader will have noted that criterion-referenced testing does not eliminate the arbitrary process of grading, since you still have a pass-fail cutoff and, wherever you set that cutoff, there is likely to be a nonsignificant difference between the scores most adjacent to it. In this kind of a grading system, however, there is nothing malicious about the arbitrary nature of the cutoff. This is because each student gets other opportunities to demonstrate that he has achieved the objectives and thus he is not stigmatized as inferior. It is also much easier for the instructor to avoid being defensive about his assessment technique, to admit to its arbitrariness and to his human frailties in assessing some other person's knowledge, and to discuss the material with the student to come to some mutual agreement as to whether further study would be desirable.

The single most important criticism of criterion-referenced assessment techniques is that you need to establish criteria or behavioral objectives for the course. For many courses this is no problem, especially if the subject matter is logically organized into sequences of material, each level of which is necessary as a prerequisite for understanding the next level. The criticism is more cogent for courses in which creative products or novel solutions to problems are the goal. In these cases, by definition, you cannot state a strict behavioral objective ahead of time. Now consider a class in which each student is pursuing a different creative project as described. Usually the student (in consultation with the instructor) establishes his goal -- what he expects to gain from the experience. How, then, should each student's progress toward that criterion, or the product that results, be evaluated?

The way it should most certainly not be evaluated is through some competitive procedure for the same reasons as given above, plus the additional reason that there is no rational way to compare performances which have different objectives. Should the student's work not be assessed? Non-assessment is reasonable only if the instructor is willing to take no credit for guiding the student's thinking, encouraging his interest, etc., in which case the instructor is superfluous to the process. The only reasonable recourse, it seems to me, is for the instructor and student to establish evaluative criteria as the project evolves. In practice this would involve student-faculty conferences in which the instructor gives the student feedback at various points in the special project. In this way the assessment process becomes an integral part of the learning-creative process, which is as it should be. Though the criteria are not stated in pre-established behavioral terms, the process is still very much criterion-referenced or, if you prefer, goal-directed with evaluative feedback from the instructor.

One final point. Many will dismiss the idea of criterion-referenced assessment because of the greater amount of work it requires to develop such a system, especially for large classes. Such procedures have been successfully developed, but they do require more work, at least initially, than either no assessment or competitive assessment techniques. However, no one ever claimed teaching to be easy. More important, excellence in teaching will continue to be in the same short supply as it is presently unless faculty members adopt some variations of criterion-referenced assessment procedures.